Investigating the Role of Microbial Proteins Found in Clinical Samples Relevant to Human Disease

Katherine Do[1,2], Subina Mehta[1], Andrew Rajczewski[1], Monica Kruk[1], Emma Leith[1], James Johnson[1], Reid Wagner[1], Pratik Jagtap[1], Timothy Griffin[1]

[1]University of Minnesota-Twin Cities, Department of Biochemistry, Molecular Biology, and Biophysics; [2]Life Sciences Summer Undergraduate Research Program

Introduction

The microbial role in early cystic fibrosis lung disease

- Obstructive lung disease is the most prominent cause of morbidity and early mortality in cystic fibrosis (CF) (1). Organisms, such as *Pseudomonas aeruginosa* and *Burkholderia cepacia*, have been identified as "traditional" pathogens that drive inflammation and structural damage in CF (2).
- Bronchoalveolar lavage fluid (BALF) has been proposed as the ideal airway sample because upper airway cultures could not accurately predict lower airway microbiota (1,3).
- Recently, anaerobic bacteria have been identified as the dominant bacterial community in BALF of asymptomatic infants with CF during their first year of life. However, it is unclear how these microbial communities may contribute to early CF lung disease in infants and young children (1).

Objective

- To determine whether microbial communities in the lower airways contribute to early CF lung disease, BALF samples were analyzed using mass spectrometry (MS)-based metaproteomics, specifically tandem MS (MS/MS) and metaproteomic bioinformatic tools on the Galaxy platform.

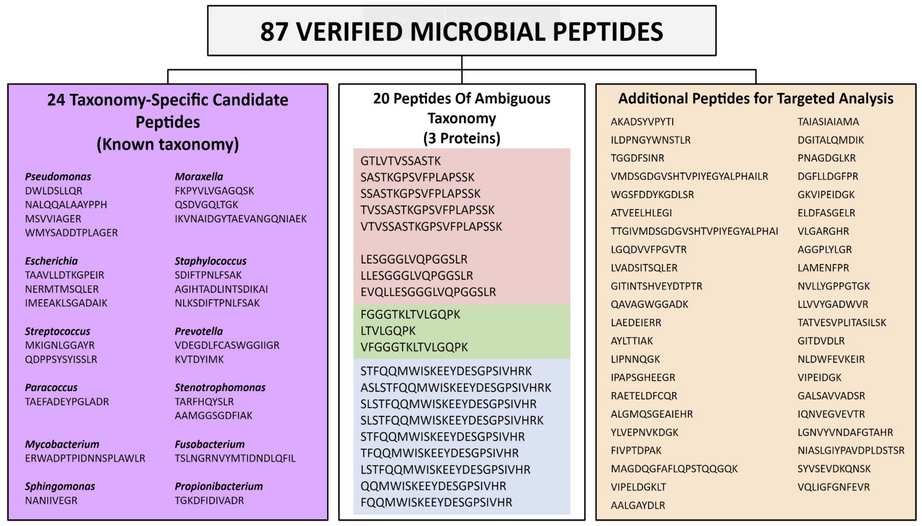


Figure 1. Breakdown of 87 Verified Microbial Peptides. From left to right: 24 taxonomy-specific peptides spanning 12 genera, 20 peptides from three proteins with ambiguous taxonomy, and additional peptides for targeted analysis, giving a total of 87 verified microbial peptides.

Results

- We identified 24 taxonomy-specific peptides, spanning 12 genera (**Figure 1**). Most were more enriched in CF samples. Some genera are (traditionally) associated with CF: *Pseudomonas, Staphylococcus, Stenotrophomonas, Prevotella, Mycobacterium* (2). "Non-traditional" genera have also been emerging: *Streptococcus, Escherichia coli* (8).
- We also identified 20 peptides of ambiguous taxonomy, but known function (**Figure 1**). Possible "non-traditional" CF pathogens include *Microbacterium* sp. AISO3, *Escherichia coli*, and potential functions include immunoglobulin (Ig)-like domains and actin-like activity.

Conclusions

- Using three different MS/MS database search programs enabled us to generate confident peptide matches with more strict verification of microbial origin. Different complementary algorithms allowed for additional filtering of possible microbial peptide sequences to retain high confidence.
- After performing Unipept and BLAST-P analysis, it was unclear whether the 20 peptides assigned to the three proteins were of microbial origin or host-origin. Ig-like domains and actin protein and/or their homologous structures found in both humans and microbes. There may also have been protein interference from the integration of database search tools (ex: match one peptide to multiple proteins).
- Here, we used pooled patient samples; the next step would be to determine whether the same peptides are detected in individual patient samples and whether there is a significant quantitative difference between CF and DC groups.

Methods

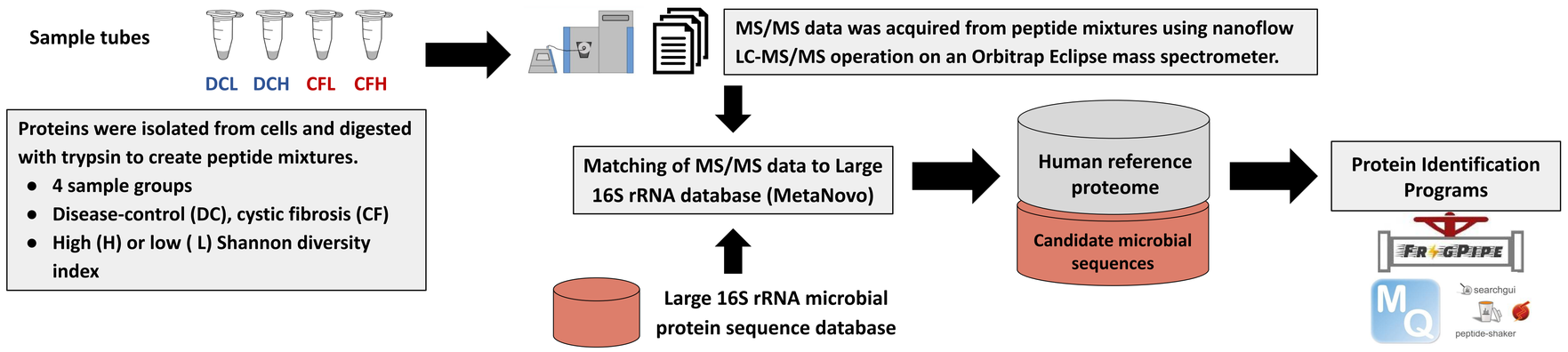
Peptides and Proteins Detection & Quantitation

Mass spectrometry offers improved microbial peptide analysis

- Mass spectrometry (MS)-based metaproteomics offers a powerful approach to analyzing microbial communities in BALF and their contributions to CF disease. Metaproteomics is the large-scale characterization of the entire complement of proteins expressed by microbiota (4,5).
- This approach provides insight into how the microbiome responds to a diseased condition. However, the abundance of human host proteins can hamper the detection of lower abundant microbial proteins in BALF.

The Galaxy-P platform for metaproteomics

- The Galaxy-P platform provides powerful tools for metaproteomic data analysis that enable users to develop and execute complex workflows that encompass data generation, peptide spectral matching, taxonomic and functional analysis (6,7).



Microbial Peptide Verification

Quantitation

- Generate abundance intensities for peptides

PepQuery

- Verify microbial peptides
- No human peptides assigned

Query Tabular

- Filter out potential contaminants
- Retain high-confidence peptide matches

Taxonomy & Functional Analysis

UniPept

- Match peptide sequences to the UniProt database
- Identify the likely genus and species

BLAST-P

- Match peptides to microbial protein sequences from the non-redundant (nr) database from National Center for Biotechnology Information (NCBI)

MEGAN6

- Taxonomic classification of peptide matches

References

1. Laguna TA, Wagner BD, Williams CB, et al. Airway Microbiota in Bronchoalveolar Lavage Fluid from Clinically Well Infants with Cystic Fibrosis. Wilson BA, ed. *PLOS ONE*. 2016;11(12):e0167649. doi:10.1371/journal.pone.0167649
2. Elborn JS. Identification and management of unusual pathogens in cystic fibrosis. *J R Soc Med*. 2008;101(1_suppl):2-5. doi:10.1258/jrsm.2008.s18002
3. Armstrong DS, Grimwood K, Carlin JB, Carzino R, Olinsky A, Phelan PD. Bronchoalveolar lavage or oropharyngeal cultures to identify lower respiratory pathogens in infants with cystic fibrosis. *Pediatr Pulmonol*. 1996;21(5):267-275. doi:10.1002/(SICI)1099-0496(199605)21:5<267::AID-PPUL1>3.0.CO;2-K
4. Batut B, Hiltemann S, Bagnacani A, et al. Community-Driven Data Analysis Training for Biology. *Cell Syst*. 2018;6(6):752-758.e1. doi:10.1016/j.cels.2018.05.012
5. Timothy J. Griffin, Pratik Jagtap, James Johnson, Clemens Blank, Subina Mehta, 2021 Metaproteomics tutorial (Galaxy Training Materials). https://training.galaxyproject.org/archive/2022-01-01/topics/proteomics/tutorials/metaproteomics/tutorial.html
6. Blank C, Easterly C, Gruening B, et al. Disseminating Metaproteomic Informatics Capabilities and Knowledge Using the Galaxy-P Framework. *Proteomes*. 2018;6(1):E7. doi:10.3390/proteomes6010007
7. Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res*. Published online April 21, 2022:gkac247. doi:10.1093/nar/gkac247
8. Edwards BD, Somayaji R, Greysson-Wong J, et al. Clinical Outcomes Associated With Escherichia coli Infections in Adults With Cystic Fibrosis: A Cohort Study. *Open Forum Infect Dis*. 2020;7(1):ofz476. doi:10.1093/ofid/ofz476

Acknowledgements

Research supported by: HLB R25 grant R25HL088728, NIH grant R01HL136499

National Heart, Lung, and Blood Institute

Medical School, University of Minnesota, Driven to Discover®